DOCUMENT RESUME

ED 183 566 TM 010 009

AUTHOR Steinmetz, Andres
TITLE The Discrepancy Evaluation Model. I. Basic Tenets of the Model.
INSTITUTION National Council on Measurement in Education, East Lansing, Mich.
PUB DATE 76
NOTE 8p.: For related document, see TM 010 007
AVAILABLE FROM National Council on Measurement in Education, 1126 16th Street N.W., Washington, DC 20036 ($1.50)
JOURNAL CIT NCME Measurement in Education: v7 n1 Win 1976

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Differences: Evaluation Criteria: *Evaluation Methods: Evaluators: Machine Repairmen: *Models: Objectives: *Performance: Program Evaluation: Staff Role: *Standards
IDENTIFIERS *Discrepancy Evaluation Model: Motorcycles

ABSTRACT

The basic principles of the discrepancy evaluation model (DEM), developed by Malcolm Provus, are presented. The three concepts which are essential to DEM are defined: (1) the standard is a description of how something should be: (2) performance measures are used to find out the actual characteristics of the object being evaluated: and (3) the comparison of standard and actual performance yields discrepancy information, which may be used to make judgments about the object's worth, based upon the differences between standard and performance. An extensive example is presented which describes the processes used by a mechanic to judge the condition of a motorcycle. The role of the evaluator in evaluating social and educational programs is then discussed in terms of the client's responsibility for setting standards, the difficulty in specifying standards, the acceptability of objectives, and the goal of program improvement. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED183566

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

NCME

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

measurement

A SERIES OF SPECIAL REPORTS OF THE NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

Winter 1978

The Discrepancy Evaluation Model

Andrés Steinmetz
President, New Measures, Inc.

ABOUT THIS REPORT

Volume 7, Numbers 1 and 2 present a departure from the traditional ME format. Both issues are devoted to the same topic, the Discrepancy Evaluation Model (DEM). Volume 7, Number 1 presents the basic principles of the model as well as an example clarifying the role of the DEM evaluator. Volume 7, Number 2 expands the discussion to include an application of the DEM to an educational program. Because of the need to introduce the readership to some of the concepts underlying the development of the model, it was felt that a two-part presentation would be in order. As with other NCME publications, ME will respond to various needs and occasionally alter its format to meet these needs.

Dr. Andres Steinmetz worked with Malcolm Provus at the Evaluation Research Center, University of Virginia, from 1971-1975. His professional experiences include many consultancies in the areas of program development and evaluation. He has published in the professional literature and has participated in many professional presentations. At present he is working as a private consultant in organizational development and program evaluation. He is president of New Measures, Inc., Charlottesville, VA, 22903.

PSR

The word evaluation is used loosely to encompass many different activities and purposes. When educators evaluate a reading program they may be referring to deciding which of several reading programs their school district should adopt; when evaluating a school bell schedule, they may mean finding out how popular the schedule is among students and faculty and what the advantages and disadvantages of several other bell schedules may be; when evaluating students they may mean administering achievement or psychological tests; and so on.

Also, the more educators stress the need for evaluation and the more it is associated with accountability and funding decisions, the more the term appears in their vocabulary. People become willing to call a lot of things evaluation when they need to show that they have done something called evaluation.

While a wide variety of activities is encompassed by the term, there is an apprehensiveness associated with it that seems to remain invariant. Evaluation suggests making judgments of worth and these judgments are generally accompanied by strong emotional reactions. The term raises apprehension that judgments will be made which will affect the

FROM THE EDITOR

Please note a change of address for editorial correspondence:

Dr. Paul S. Redelheim
Project MORE, Director of Research
George Peabody College for Teachers
Box 318
Nashville, TN 37203

This change is effective immediately.

PSR

TM010 009

social and/or professional status of people, their career plans, their self esteem, the scope of their authority, and so on.

What all this amounts to is that when one is called upon to do evaluation, it is usually hard to escape first doing battle with a lot of expectations people have about what is going to happen. To work effectively, a practitioner is forced to clarify his/her position relative to all these expectations. The Discrepancy Evaluation Model (DEM)* represents an assembly of ideas and procedures arising out of attempts to respond constructively to such expectations. It represents a scheme with which to respond to the challenges presented by the difficult task of evaluating educational programs.

I. BASIC TENETS OF THE MODEL

The Concepts of Standard, Performance, and Discrepancy

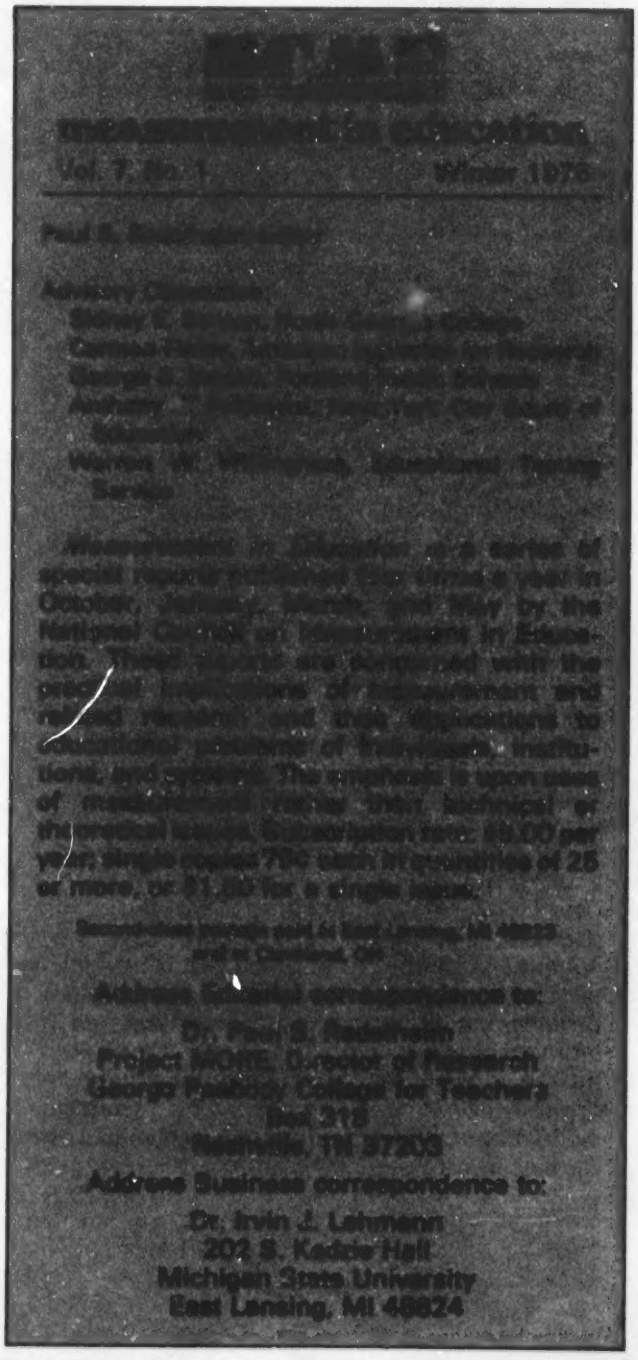
In order to evaluate something, we inevitably make comparisons. More specifically, we say that to evaluate a given object (whether a person, a motorcycle, or a program) it must be compared to a Standard. By a Standard we mean a list, description, or representation of the qualities or characteristics the object should possess. In other words, a description of how something *should be* is called the Standard (**S**).

Once we are clear about how things should be, we can proceed to find out whether they *actually are* that way. When we are engaged in finding out the actual characteristics of the object to be evaluated, we are taking Performance measures (**P**). Thus, evaluation is a matter of comparing **S** against **P**.

There is another term involved in the *comparison* between **S** and **P**. We say that the comparison yields Discrepancy (**D**) information and thus we can speak of evaluation as being a matter of making judgments about the worth or adequacy of an object based upon **D** information between **S** and **P**.

The concepts of **S**, **P**, and **D** surface quite naturally whenever, under the name of evaluation, one wants to judge the adequacy or worth of something. Suppose, for example, that you want to purchase a motorcycle but are uncertain whether the specific one you are considering is in good mechanical condition and you therefore arrange to have a mechanic examine it. We can use the concepts of **S**, **P**, and **D** in order to describe what the mechanic will do pursuant to your request to find out whether the motorcycle is in good mechanical condition. Essentially, the mechanic will take certain **P** measures and compare these to an **S**. The **D** information generated in making the comparison will somehow aggregate into a judgment about whether the motorcycle is or is not in good mechanical condition.

The mechanic has some ideas about how the motorcycle *should be* functioning when it functions adequately and he/she will proceed to test these out. For example, he/she may refer to the motorcycle specifications manual to find out what compression the pistons should generate and that informa-



*The DEM was first put forth by Malcolm Provus; see *Discrepancy Evaluation*, McCutchan, 1971. It was further developed at the Evaluation Research Center in 1971-1975 by a team of people, including the author, led by Malcolm Provus. Different versions of the DEM have arisen. The views presented here are those of the author.

tion will become part of the **S**. Then, as a **P** measure, he/she can find out whether the pistons do in fact generate the compression specified. He/She may also listen to the way the motorcycle is idling (**P**) and compare that to his/her experience in order to decide whether the engine sounds the way *it should* sound (**S**). Or, he/she may refer to both his/her experience and the specifications manual (**S**) in order to generate **D** information about the actual condition (**P**) of the brakes. As another **P** information collecting strategy, the mechanic is also likely to drive the motorcycle and compare how it feels and sounds with how he/she thinks the motorcycle *should* feel and sound.

Thus, to find out whether the motorcycle is in good mechanical condition, the mechanic will do certain things like measure compression, test the brakes, examine the spark plugs, etc., all of which represent gathering **P** information. Of course, the mechanic will probably restrict himself/herself to collecting **P** measures on a limited number of dimensions according to time available, the price he/she has agreed on with you, his/her experience about what is important to look at, and, also, the availability of an **S** governing what he's/she's looking at. Note also that **S** will vary in specificity and will be a mixture of the mechanic's experience and the operating and engineering specifications of the motorcycle. The **D** information he/she generates by comparing **S** and **P** will become a basis for the conclusions he/she submits to you. In his/her conclusions he/she is likely to pass a judgment by saying that the motorcycle is or is not in good condition. He/she will probably substantiate his/her conclusion by referring to some of his/her findings. He/She might add things like, "the piston rings are worn," or "it needs a new clutch" and thus roll into one phrase a **P** statement, an **S**, **D** information between the two, and a conclusion indicating what to do about it. Furthermore, knowing that you are trying to decide whether or not to buy the motorcycle, he/she is likely to make a recommendation such as, "it's OK for the price," or "I wouldn't buy it."

In similar fashion, the **S**, **P** and **D** concepts can be shown to underlie the making of any judgments of adequacy or worth. More than that, they seem to underlie any cybernetic process and much of human behavior. Under the DEM, however, the important thing is how these concepts are applied. To discuss how they are applied, I would first like to make some summary observations about the work of the mechanic in the example considered above. Then, I will consider the role of a DEM evaluator by elaborating a bit further on the same motorcycle example. At that point we will be in a position to understand the application of the DEM and can go on to apply it to something a bit more complex than a motorcycle — an educational program.

Summary Observations

a. The person you consulted to determine whether the motorcycle is in good mechanical condition is considered by you to be an expert in motorcycle mechanics.

b. Both the **S** and the specific characteristics of the motorcycle to be examined (the specific **P** information to be gathered) were selected and determined by the mechanic. In particular, the sources of the **S** were the mechanic's experience and knowledge, and the manufacturer's specifications.

c. Much of the **S** was left implicit. You might ask to have the **S** applied verbalized or explained to you, but the custom governing the exchange between you, the client, and the mechanic (as expert and as evaluator) tends to keep that sort of conversation to a minimum and on a relatively superficial level. And, whatever conversations might be held about **S** are usually jargon loaded and assume knowledge over the very conditions or phenomena which, as client, you don't have, and which led to your turning to someone else in the first place. Thus, the specific **S** brought to bear on certain performance information may remain unknown to you and to some extent, because they are not articulated by him/her, less than consciously known to the mechanic. For example, in examing the condition of the spark plugs the mechanic may notice that they are a brand which he/she considers inferior to another brand. That, in itself, may tend to make him/her more willing to consider their life exhausted than he/she might otherwise be. And the influence of that bias in his/her judgment is something the mechanic may not be ready to acknowledge.

d. You are not likely to see the specific performance information obtained. The mechanic is likely to report to you an overall judgment about the mechanical condition of the motorcycle, elaborate a bit on some of his/her findings, and respond to some questions you might have. He/She is unlikely, however, to itemize **P** and **D** information for you or to be explicit about how he/she aggregated the **D** information to arrive at his/her judgment. He/She may also recommend to you a course of action — repair, price negotiation, etc.

The Role of the Evaluator

We have seen how what the mechanic was asked to do can be discussed in terms of **S**, **P** and **D**. Thus, we can say that the mechanic was evaluating the mechanical adequacy of the motorcycle. But we would not say that he/she was applying the DEM even though we can describe what he/she did in terms of **S**, **P** and **D** because the critical thing about the DEM is the *manner* in which these concepts are applied. The crucial thing rests in the role relationship which is assumed by the evaluator vis-a-vis the client. In particular, the DEM evaluator would neither set **S** nor judge the comparisons made between **S** and **P**, though he/she would normally collect **P**. Instead, he/she would assist the client to do these things for himself/herself.

Let's suppose, to explore this role relationship, that you come to me, a DEM evaluator, but neither a motorcycle expert nor mechanic, and ask me to help you with your larger problem — namely, to evaluate a specific motorcycle with the aim of deciding whether or not to purchase it. As a DEM evaluator, the first thing I would be concerned about is the existence of an **S**. I would want to know from you what you are looking for in a motorcycle, what characteristics or qualities you feel that motorcycle *should* possess. If my turning to you for an **S** seems a little odd at this point it may be because we have some different ideas about and expectations from an evaluation. Remember that this role characteristic of the evaluator is chosen in order to permit constructive response to technical, political, organizational, and emotional problems encountered in the applied situation.

As a DEM evaluator, I would be seeking a model representing the kind of motorcycle you are looking for which can then be used as the **S** against which to compare any particular motorcycle. This stands in contrast to summary observations (a) and (b) above which noted how the mechanic (as evaluator) was considered an expert authority and the source of the **S** governing mechanical functioning. I might begin by helping you make a list of the characteristics or qualities you value or find desirable as shown in Figure 1.

Characteristics the motor cycle should possess:

- *Cost:* should not cost over $800
- *Power:* should be able to cruise at 60 mph
- *Stability:* should be large and heavy enough to stay on the road
- *Noise:* should be quiet
- *Appearance:* should have the classic "World War I look"
- *Mechanical Condition:* should be in good condition and not presently need repairs

Figure 1
First Attempt at a Standard

While this first attempt at an **S** gives an idea of the kinds of characteristics you feel the desired motorcycle *should* possess, I would still need further guidance from you before I could collect **P** information that would be useful to you. I would ask you to formulate some questions — evaluation questions — which you would want answered relative to each characteristic making up the **S**. These would be questions which ask directly whether the quality, condition, or characteristic desired and specified by the **S** obtains in reality. Let's take cost and power, for example. In the case of cost, an evaluation question might be: What does this specific motorcycle cost? In getting ready to answer this question, I would make myself a little work plan, something like that shown in Figure 2, which I would review with you.

(See Figure 2 at bottom of page.)

It would be clear that when I get the answer to question one in dollars, it will be easy for you to determine whether or not the **S** is met, since you have said it should not cost over $800. Let's consider power, however. As shown in Figure 2, an evaluation question here might be: Does this motorcycle cruise at 60 mph? In thinking through the elements of the evaluation workplan here, however, I would be faced right away with a basic problem and I would turn to you for clarification: I can answer the question in many different ways and still meet your Standard. For example, I could ride down a long incline and maintain 60 mph, or drive a straightaway with no passengers or wind and maintain 60 mph for one-half hour, or I could find that the engineering specifications say the motorcycle will cruise at 65 with one passenger, or so on. So, I would think through these difficulties with you and urge you to set an **S** that restricted more severely the number of different conditions and answers that would satisfy it. We might thus end up with a more specific **S** as shown in Figure 3.

(See Figure 3 at top of page 5.)

Let's now turn to another aspect of the **S** mentioned in Figure 1, i.e., stability. We can see right away that the same problem with the **S** reappears here. You claim a relationship between size, weight and stability but if we don't know what it is then simply finding out the size and weight of a given motorcycle will not let you know whether your

Evaluation Question	Standard	Source of Information	Instrument	Data Collector	Date Info is Needed
1. How much does this motorcycle cost?	The motorcycle should not cost more than $800	Seller of motorcycle	Interview	Evaluator	Next Friday
2. Does this motorcycle cruise at 60 mph?	Should be able to cruise at 60 mph				

Figure 2
Elements of an Evaluation Workplan

Evaluation Question	Standard	Source of Information	Instrument	Data Collection	Date Info Needed
1. How much does this motorcycle cost?	The motorcycle should not cost more than $800	Seller of Motorcycle	Interview	Evaluator	Next Friday
2. Does this motorcycle cruise at 60 mph?	It should maintain 60 mph with 2 passengers on a straightaway all day.	Engineering specifications	Review of engineering specs	Evaluator	Next Friday
	It should maintain 60 mph with 2 adults up the mile-long hill on Rt. 629	Motorcycle	Road test	Evaluator	Next Friday

Figure 3
Elements of an Evaluation Workplan

S for stability will have been met. Again, I would urge you to decide what you will consider adequate stability and to agree with me on an appropriate way of taking performance measures. If you could reach no conclusion on your own, I would help you think through a number of different options. You could consult motorcycle engineers for constellations of variables and conditions that might define stability; you could launch a research project yourself aimed at defining stability and the factors involved, etc. You could also decide to remove the whole matter of stability from your **S**. While I would facilitate an **S**-setting process and seek to confront you with decisions you would need to make in order to have an **S** available, I would not get involved in the work or decisions involved in creating the **S** itself. Thus, for example, I would not carry out the research project that would build the **S** — unless I completely changed my contract with you and it was made clear I would no longer be an evaluator. To me, evaluation would presume the existence of the **S** and would entail merely looking at a specific object or event to see whether pertinent characteristics or conditions are present.

Under most circumstances, I would not even collect any data unless the pertinent **S** was explicitly stated. Otherwise, you, as client, would be left open to the possibility that I, and not you, would make the ultimate judgment of adequacy in the evaluation. Suppose you decide that the research questions concerning stability are too time consuming and expensive given that you want to reach a decision within a brief time period. Thus, you are in the position of remaining interested in a certain quality (stability) yet find yourself without an **S** for it. This certainly is a common enough situation in most daily affairs and there is a popular method for dealing with it: leave the **S** unexplicated and decide on the adequacy of the **P** information as you collect it. Applying this method to our example, one could drive the motorcycle and draw some conclusion about how "stable" one feels — which involves conjuring up an **S** based on the immediate experience. One would be contrasting the immediate experience to an ideal implicit model of stability. One could also make a comparative judgment by riding a number of motorcycles and comparing the feelings of "stability" involved. Either way, one collects some psychomotor knowledge about the stability one desires, formulates an **S**, and generates **D** information while test driving. However, if this latter route were chosen in order to obtain **P** measures on stability, then *you,* as client, would have to do the test driving. You are the one, after all, who is primarily interested in knowing whether a specific motorcycle meets your **S**. If the **S** is left unobservable and if I, as evaluator, do the test driving, the matter of stability would end up being judged against *my* (implicit) **S** and not yours.

Review

While the work of both the mechanic and the DEM evaluator can be described in terms of the **S, P** and **D** concepts and thus be called evaluation, there are important differences in the way each discharged his/her role. These differences can be summarized in terms of the relationship of the mechanic and the evaluator to the **S** involved. The mechanic was the source of **S**, selected **S**, defined the **P** measures to be made, the procedures involved, and also collected the **P** information. He/She then compared **S** and **P** in each case and formed an overall judgment concerning the mechanical adequacy of the motorcycle based on the **D** information generated. He/She was also not particularly concerned with making **S** explicit nor with presenting **P** in any great detail. At least, he/she did so only as it seemed necessary to make his/her conclusions plausible and convincing to the client, or in answer to specific questions.

In contrast, the DEM evaluator approached the problem by helping the client articulate the dimensions involved in **S** by making it clear that the responsibility for deciding what **S** *should be* rests with the client. The evaluator also made it clear that the client had to specify the kind of evidence that would be an acceptable index of the **S** as well as

what would be considered criterion performance. Moreover, the significance attached to all discrepancies found, and thus, the overall judgment of adequacy, was also left to the client. The DEM evaluator was thus the facilitator of a process. The actual evaluator, in the sense of making the judgment of worth, was the client.

To ensure that the client would be in the position of making the judgment of worth, both the **S** and **P** information to be collected had to reach a certain level of specificity independent of the personality of the DEM evaluator. This is one thing that makes the role of the evaluator tricky. While the DEM evaluator would generally consider it part of his/her responsibility to collect **P** information, he/she would not do so in cases where it would evidently compromise the client's ability to compare **S** and **P** and thus give **D** significance. The example involving stability was a case in point. We saw that by leaving **S** embedded in personal experience, the very definition and collection of **P** tacitly set **S**. Therefore, in that instance, the client was asked to collect **P** himself/herself.

There are two aspects of the posture of the DEM evaluator which require further comment. They have to do with the interest of distinguishing clearly between the acts of setting **S** and determining whether **S** has been met. First, as already noted, setting **S** is the responsibility of the client but facilitating the process is the responsibility of the evaluator. If the client is unable to formulate a pertinent **S**, then he/she can undertake whatever activity necessary to create it, which may involve consulting experts or launching research projects. He/She may also engage an expert to do the "evaluation" for him/her. This was the situation in the example above where a mechanic was engaged to judge the mechanical condition of the motorcycle because the client did not feel he/she had the expertise to do it himself/herself. Yet, because of the way in which this "evaluation" was performed, or, more exactly, because of the client's relation to it, we would not consider it a DEM evaluation. The major reason for this is that the client is not expressly setting **S** and this would be considered pre-empting his/her decision-making role and responsibility. It may be objected that the client is still free to accept or reject whatever the expert ends up recommending. That is, of course, true but the point is not so much that the client end up making the final decisions as it is that he/she expand his/her awareness of the raw ingredients that go into making the decision. Choosing to accept or reject a formed judgment is different from being a party to the making of that judgment.

The second matter is a variation of the difficulty often encountered in specifying **S**. One can say that leaving the definition of something like stability to an unexplicated feeling derived from test driving is not "objective," is not "scientific," or does not provide the evaluator with an "operational" or "observable" definition. And not being "objective" or "scientific" in this sense is generally shunned. I prefer, however, not to talk about the issue in this way. I think it is better to discuss the issue in terms of roles and responsibilities. The client is the one who has to live with the choice made. He/She is the one who has to take responsibility for the evaluation and the decisions resulting from it. The specific data or performance information one responds to in com-

REPORTS AVAILABLE

Back issues of *Measurement in Education* are available at 50¢ each in quantities of 25 or more for a single issue.

Vol. 1, No. 1	*Helping Teachers Use Tests* by Robert L. Thorndike
No. 2	*Interpreting Achievement Profiles — Uses and Warnings* by Eric F. Gardner
No. 3	*Mastery Learning and Mastery Testing* by Samuel T. Mayo
No. 4	*On Reporting Test Results to Community Groups* by Alden W. Badal & Edwin P. Larsen
Vol. 2, No. 1	*National Assessment Says* by Frank B. Womer
No. 2	*The PLAN System for Individualizing Education* by John C. Flanagan
No. 3	*Measurement Aspects of Performance Contracting* by Richard E. Schutz
No. 4	*The History of Grading Practices* by Louise Witmer Cureton
Vol. 3, No. 1	*Using Your Achievement Test Score Reports* by Edwin Gary Joselyn & Jack C. Merwin
No. 2	*An Item Analysis Service for Teachers* by Willard G. Warrington
No. 3	*On the Reliability of Ratings of Essay Examinations* by William E. Coffman
No. 4	*Criterion-Referenced Testing in the Classroom* by Peter W. Airasian and George F. Madaus
Vo. 4, No. 1	*Goals and Objectives in Planning and Evaluation: A Second Generation* by Victor W. Doherty and Walter E. Hathaway
No. 2	*Career Maturity* by John O. Crites
No. 3	*Assessing Educational Achievement in the Affective Domain* by Ralph W. Tyler
No. 4	*The National Test-Equating Study in Reading (The Anchor Test Study)* by Richard M. Jaeger
Vol. 5, No. 1	*The Tangled Web* by Fred F. Harcleroad
No. 2	*A Moratorium? What Kind?* by William E. Coffman
No. 3	*Evaluators, Educators, and the Publics: A Detente?* by William A. Mehrens
No. 4	*Shall We Get Rid of Grades?* by Robert L. Ebel
Vol. 6, No. 1	*On Evaluating a Project: Some Practical Suggestions* by John W. Wick
No. 2	*Measuring Reading Achievement: A Case for Criterion — Referenced Testing and Accountability* by S. Jay Samuels and Glenace E. Edwall
No. 3	*Security in a Citywide Testing Program* by Anthony J. Polemeni
No. 4	*Dear Mama: Why don't they love me anymore?* by Dr. Thomas J. Fitzgibbon

paring **P** and **S** is certainly a matter of one's belief structure and one's preferred way of relating to the world. If nothing other than empirical or "scientific" data will do, then certainly one could proceed to construct an empirical definition of "stability." But I don't think that is automatically the best or most appropriate way to proceed. A client may find his/her own personal judgment based on "feel" or unexplicated criteria satisfactory, and if *he/she* does, I will too. That doesn't mean that I would not try to explore with the client other alternatives or the consequences of doing things in different ways. It also doesn't mean that I would accept any evaluation contract whatsoever. It means that the credibility of evidence is a function of one's beliefs and that quantitative objective data is not necessarily the thing to strive for.

An interesting corollary here is that, in the case of program evaluation, absolutely any program objective is an adequate objective as far as the DEM evaluator is concerned. There is no need to insist on behavioral or other kinds of objectives. The role description already provided emphasizes freedom to set **S** as seems desirable and pertinent to the client who carries the responsibility for the program. Rules for expressing criterion performance, in particular, are not necessarity deduced from a certain methodological orientation or logical framework. Acceptable **S** and **P** are seen rather as a function of the set of agreements and beliefs that make up the world of the client.

These role characteristics may be unwieldly when a layperson wants to evaluate a motorcycle but they are essential to the comprehensive and useful evaluation of something like an educational or social service program. This is because programs represent organized human activity and, as such, always represent normative states of affairs. For here, **Ss** ultimately are issues in social and moral philosophy, and, immediately, ideological and political matters. The position taken here is that **S** therefore not be left to experts dissociated from the responsibilities of program operation and management. For program staff to choose and commit themselves to the **S** pertinent to their clients' contexts and personalities is fully as important as having an **S** that is rooted in abstract empirical generalizations (i.e., expert knowledge).

Finally, it should be evident that setting the role of the evaluator in this way formalizes the common human evaluative activity that is associated with any deliberate act. When applied to a program, DEM evaluation refers to making explicit the procedures and norms governing the **SPD** cycles that make up planning, implementation, and review activities. Thus, DEM program evaluation is aimed at program improvement. With the client in control of **S** and guiding the collection of **P**, **D** information can be used to keep action flexible, responsive, and informed. Since **D** is the result of comparing **S** and **P**, we can reduce **D** by changing **S** or changing **P**. Changing **S** involves program redesign, perhaps changing basic objectives or activities. To change only **P** in a program requires that management exert greater control over operations.